Reliability and validity of the Meaningfulness of Songwriting Scale with university students taking a popular songwriting class

Abstract

Background: Songwriting is a fundamental and universally accessible aspect of music making. It is also emerging as a popular activity for use in therapeutic contexts suggesting that the songwriting process can be meaningful irrespective of musical experience. Our study tested the properties of the Meaningfulness of Songwriting Scale (MSS) with university students studying songwriting, to determine its validity and reliability in measuring the extent of meaning derived from the process. **Method:** Participants completed the MSS, Short State Flow Scale, and Rosenberg Self-Esteem Scale immediately following their creation and performance of an original song. Seven days later, participants completed the MSS a second time. **Results:** Analyses were performed and findings indicated that the measure has good content validity, strong internal consistency, acceptable test-retest reliability, limits of agreement, measurement error and discriminant validity. Convergent validity was weaker than the other measurement property results but still a produced a near-significant result. **Conclusions:** This study provides evidence that the measure has adequate measurement properties for use with university students and maybe utilized to further our understanding of the process and outcomes of this emerging, and potentially highly beneficial, music therapy strategy. Future research should seek to further investigate the psychometric properties of these instruments across different populations and contexts.

Keywords: songwriting, meaning, measurement properties, music therapy, flow, self-esteem

Introduction

Songwriting is a key aspect of musical engagement and while not everybody who plays or listens to music is involved in songwriting, it is a universally accessible process. It is also a popular art form that many people engage in either as a professional, amateur, for personal growth, or within therapy contexts (Baker, 2015a). The existence of songwriting competitions – International Songwriting Competition, John Lennon Songwriting Contest, Australian Songwriting Contest, Song of the Year Songwriting Contest, Vanda and Yong Songwriting Contest, to name just a few, highlights the popularity and significance of songwriting. Certainly within a therapeutic context, songwriting has emerged as a powerful approach to assist people with a range of psychosocial, emotional, cognitive, and communicative challenges to express multicultural and individual identity, adjust to life changes, and cope with life's challenges (e.g. Baker, Wigram, Stott, & McFerran, 2008/2009).

A number of efficacy studies with people with mental illness, adolescent and adult cancer patients, people with dementia, people with spinal cord injury, and people with acquired brain injury illustrate songwriting's utility to impact a range of wellbeing indicators. These include satisfaction with life, flourishing, or quality of life (Baker,

Rickard, Tamplin, & Roddy, 2015; Grocke, Bloch, & Castle, 2009; O'Brien, 2014), levels of depression (Baker et al., 2015; Silverman, 2011), motivation for treatment (Silverman, 2012), working alliance, (e.g., Silverman, 2011), coping (Robb et al., 2014; Silverman, 2011), physical wellbeing (O'Brien, 2014), and self-concept (Baker et al., 2015). However, more outcome studies are needed to identify which aspects of wellbeing are most positively impacted by songwriting, which populations are most responsive, length, and frequency of songwriting sessions lead to the best outcomes, and what therapeutic orientation songwriting is practiced, is (if at all) the most effective (Baker, 2015b). Currently, no studies have measured the impact of songwriting on non-clinical populations' wellbeing.

To design songwriting programs for clinical and non-clinical populations (e.g. primary, secondary, or university students, or artist-led songwriting programs [Imms, Jeanneret, & Stevens-Ballenger, 2011]), an understanding of the mechanisms of change activated by the songwriting process is needed (Baker, 2016). Recent research has begun to address this need to understand how and why songwriting impacts wellbeing. In 2013, Baker and MacDonald (2013a/b) undertook qualitative and quantitative research with University students and healthy retirees to explore the experience of songwriting and the meaning of the songs created in a single session facilitated by a trained music therapist. In

this study, participants were asked to create three original songs about a positive experience, a negative experience, and a neutral experience, They were then asked to rate their experience of flow and their experience of the meaning of the songwriting process and product in a questionnaire constructed specifically for the study. Participants were also interviewed about their experience to get a more nuanced sense of the meaningfulness of the experience. Quantitative data indicated that songwriting fostered a strong sense of flow and meaning, and that these were highly correlated and formed a predictive relationship.

Baker and MacDonald reviewed the qualitative (Baker & MacDonald, 2013a) and quantitative findings (Baker & MacDonald, 2013b) and developed the Meaningfulness of Songwriting Scale (MSS), which aimed to measure the extent of meaningfulness derived from the songwriting process and product. The purpose of this measure was for future use in research studies so that research could correlate outcomes with the degree of meaning experienced by the participants. The measure has undertaken preliminary property testing with a sample of 39 people who have acute psychiatric illness and a sample of 108 people undergoing detoxification from drug addiction (Baker, Silverman, & MacDonald, 2016). Participants were part of a single session group songwriting experience (small groups of approximately 8 participants per group) where groups created songs that focused on hope (acute psychiatric group) or sobriety (detoxification group). Twelve-bar blues songs were

created whereby lyrics focused on why participants wanted to recover from their illness or be sober and themes that emerged focused on family, friends, health, vocational aspirations, and happiness (Silverman, Baker, & MacDonald, 2016). Findings indicated that the measure has good face validity, strong internal consistency ($\alpha = 0.98$, acute psychiatric group and $\alpha = 0.96$, detoxification group), acceptable test-retest reliability ($r = 0.93$, acute psychiatric group, and $r = 0.89$, detoxification group), and convergent validity (acute group was $r = 0.68$, $p<0.001$, and detoxification group was $r = 0.56$, $p<0.001$).

Baker et al. (2015) used the MMS with 5 adult in-patients with acquired brain injury and 5 adult in-patients with spinal cord injury and correlated their scores with a range of wellbeing issues. The songwriting protocol involved participants creating three songs over 12 twice-weekly sessions – one about their past self, one about their present self, and one about an imagined future self – and collected data on levels of self-concept, depression, anxiety, emotion regulation, affect, satisfaction with life and flourishing at pre, midpoint, and post-test. The MSS was completed 3 times for each participant – once at the end of each of the composed songs. Correlations showed that deriving high levels of meaning was associated with increased negative affect, increased anxiety, and reduced emotional

suppression. Importantly, the findings support the use of the MSS in understanding the mechanisms activated during songwriting processes. To strengthen the validity of future research, testing the reliability and validity of the properties of the MSS is needed. This study aimed to provide data with a sample of healthy university students so that future research could compare responses to songwriting with a non-clinical sample of participants. Specifically we aimed to answer the following research questions:

1) Internal consistency: Are the MSS items interrelated and interchangeable and therefore does the MSS demonstrate sufficient internal consistency?
2) Test-retest reliability and measurement error: Does the MSS have acceptable test-retest reliability and measurement error when participants complete the measure on two occasions separated by 1-week?
3) Convergent validity: Is the construct of meaningfulness tested in the MSS related but still distinctly different to the construct of flow as measured in the Short Flow State Scale (SFS)?
4) Discriminant Validity: Is the construct of meaningfulness tested in the MSS sufficiently unrelated to the construct of self-esteem as measured in the Rosenberg Self-Esteem Measure?

Method

Participants and Context.

The data used to test the properties of the MSS were collected during 3 separate songwriting classes held once per semester for three consecutive semesters at a leading University in XXXXX. This class, an elective that can be taken by any students enrolled in any of the Bachelor degrees offered at the university, introduces the students to the basics of songwriting for the commercial music industry. No pre-requisite skills are required and many of the students have little or no formal musical training. The "popular songwriting class" combines 12 formal lectures and practical workshops in a large group format and is designed to enable students to understand the fundamental principles of songwriting. Classes were 2 hours in duration and held once weekly over 13 consecutive weeks with a one-week mid-semester break occurring at week 6 or 7. Learning activities and assessment include the completion and submission of an original popular song that is based on songwriting and commercial music concepts presented during lectures and workshops. Students worked on their songs within the workshops and during their own individual study time. The song was assessed by the lecturer (not one of the researchers) and formed 60% of the assessment. Part of the class requires students to present their song to the class.

The project was reviewed by the University of XXXXX Human Research Ethics Committee and classed as minimal risk (approval number 1339755.1). All participants were required to give informed consent to participate. Seventy-one students enrolled in the class across the three semesters and of these, 28 students consented to participate in the research. Demographic data were not collected but the typical cohort was 18 to 23 years of age, there tended to be twice as many male compared with female students enrolled in the class, and students were enrolled as first year, second year, or third year Bachelor degree students.

Procedures.

Once recruited to the study, participants were asked to complete the MMS (testing MSS's internal consistency), the Short State Flow Scale (SSFS, Martin & Jackson, 2008) to assess flow (testing for convergent validity), and the Rosenberg Self-Esteem Scale (Rosenberg, 1965) to assess self-esteem (testing for discriminant validity), immediately after their performance of their original song within the class context. Seven days later, the participants completed the MSS a second time within class time (testing for measurement error and test-retest reliability). The last group of students to perform their songs in week 12, did not have an opportunity to complete the second MSS in class time and as a result some data were missing (see results).

Measures.

Meaningfulness of Songwriting Scale.

The MSS (Baker et al., 2016) comprises 21-items designed to measure the degree of meaning across three dimensions: affective meaning, cognitive meaning, and relational meaning. These dimensions are explored through items that reflect 11 domains of meaningfulness of a songwriting process and the song product: enjoyment, discovery/self reflection, arousal of emotions, creativity, engagement, challenge, understanding context, associations, achievement, personal value, and identity (Baker et al. 2016). Six items relate to affective meaning, 10 items relate to cognitive meaning, two items capture both affective and cognitive meaning, one item reflects relational meaning, and two items capture all dimensions of meaning (Baker et al., 2016). The development of scale including testing face validity is described in detail elsewhere (Baker et al., 2016).

In the MSS used in Baker et al. (2016), all items were framed in the positive direction and measured using a 5-point Likert scale from Strongly Disagree to Strongly Agree. Based on response trends from that study, this study reworded six items (item 1, 6, 7, 9, 11, and 14) so they were structured in the negative direction (for example, item 1 was changed from “Songwriting was an enjoyable process” to “songwriting was not an enjoyable process”. Scoring for these items was reversed. Total scores range from 21 to 105 with

higher scores indicative of greater meaning derived from the process and song product (Baker & MacDonald, 2013a/b).

The Short State Flow Scale (SSFS, Martin & Jackson, 2008) was selected to test convergent validity because it measures a different but related phenomenon to meaning and demonstrated convergent validity in a previous study testing the properties of the MSS (Baker et al., 2016). The measure contains 9-items that capture the respondent's perceived state of flow in response to a specific context. The scale's items reflect nine dimensions of flow (Jackson & Csikszentmihalyi, 1999), specifically challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration, sense of control, loss of self-consciousness, transformation of time, and autotelic experience. Items are measured along a 5-point Likert scale with scores ranging from 9 (no flow experience) to 45 (ultimate flow experience). Previous studies have established the SSFS has good construct validity across several settings (work, sport and music; Martin & Jackson, 2008) and demonstrates acceptable reliability (Cronbach's $\alpha = .82$). Internal validity for flow in music ($\chi 2 = 44.11$) was higher than for work ($\chi 2 = 136.78$) and sport ($\chi 2 = 112.38$) and

external validity was also high for music ($\chi 2 = 4056.76$), work ($\chi 2 = 6088.56$), and sport ($\chi 2 = 4479.03$).

The Rosenberg Self-Esteem Scale (RSES, Rosenberg, 1965) measures individual levels of self-esteem and was chosen to determine discriminant validity of the MSS. It contains 10 items that ask the respondent to reflect on his or her beliefs about the self as a means to measure levels of self-esteem. Items are scored on a four-point Likert scale ranging from 'strongly agree' (scored with a 3) to 'strongly disagree' (scored with a zero). Five items (items 2, 5, 6, 8, and 9) are reversed scored. The RSES has been used extensively in music intervention studies (e.g. Choi, Lee, & Lee, 2010; Porter et al., 2012; Wood, Ivery, Donovan, & Lambin, 2013). It has been validated using a large sample of high school students (Rosenberg, 1965), and has since been validated for use with adults in psychiatric care and the general population (Blascovich & Tomaka, 1991). Measurement properties testing indicates it has high test-retest reliability with correlations ranging from 0.82–0.88 and Cronbach's α ranged from 0.77 to 0.88 (Blascovich & Tomaka, 1991).

Data Analysis.

Using calculations proposed by Walter, Eliasziw, and Donner (1998) where n = 2 (test-retest reliability), $\rho_0 = 0.8$, $\rho_1 = 0.9$, $\alpha = 0.50$, and $\beta = 0.20$, with an anticipated dropout rate of 10%, we planned to recruit 50 participants. This is similar to other sample sizes used in testing the measurement properties of scales about responsive to various music conditions (e.g. Clark, Baker, Peiris, Shoebridge, & Taylor, 2015; Hald, Baker, & Ridder, 2015; Magee, Siegert, Taylor, Daveson, & Lenton-Smith, 2016) and comparable with the acute psychiatric sample in an earlier validation of the MSS (Baker et al., 2016).

Internal Consistency: Cronbach's coefficient alpha (α) was applied to all the available test and retest MSS measurements to determine the strength of correlation between the 21 items on the MSS. The mean and standard deviation between each of the 21 items and the total score were calculated to determine the strength of inter-item correlations. Internal consistency was considered satisfactory if item correlation was >0.70 (Shrout & Fleiss, 1979).

Test-retest Reliability: Pearson's product-moment coefficient (*r*) was used to examine the strength of correlation between the total scores of the first MSS and the second MSS, which was completed seven days later. Currently no universal consensus as to what constitutes a sufficient correlation coefficient, however, in this study, any figure > 0.7 was

considered acceptable and figures > 0.8 were considered to indicate good reliability (Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013).

Measurement Error: The distribution of observations for agreement between the difference scores at measurements 1 and 2 was calculated using limits of agreement (Portney & Watkins, 2009). Range of error for any individual participant was determined by plotting the difference between the two measurements (1 and 2) against the mean score for each participant (Bland-Altman plot, Bland & Altman, 1986). T-tests and sign-tests calculations were used to determine whether there were significant mean differences between the test and retest measurements.

Convergent and Discriminant Validity: Pearson's product-moment coefficient (r) was used to examine the strength of correlation between the MSS and the SSFS with a moderate value of $r = 0.4 – 0.6$ indicating acceptable correlation and evidence of convergent validity. Scores for the correlation between MSS and RSES of $r < 0.2$ were indicative of discriminant validity.

Missing data: Missing data were treated in the same way as per Baker et al. (2016). When only one item was missing from the post-session or 7-day retest MSS (responded to 20 of 21 items), the SSFS (8 of 9 items), and the RESE (9 of 10 items), an average of the

total items was calculated and missing values replaced with the average response for that participant. Where two or more items per scale had missing data, the total score was not included in the pooled analyses.

Results

Missing Data.

There were a number of participants who had missing data. For the post-session MSS, two of the 30 participants answered 20 of 21 items; the other 28 participants fully completed the measure. Two of the 30 students did not submit the SSFS, and one participant answered 8 of the 9 items. All other participants answered all 9 items. Three participants did not submit their RSES for analysis and of the remaining 27 participants, only one answered 9 of the 10 items. All other participants fully completed the measure. As participants did not provide contact details, 12 of the 30 participants who performed their song in the last week of their class and did not return their 7-day MSS retest. Therefore the test-retest analysis is based on only 18 participants. Among these 18 participants, one answered 20 out of 21 items; the other 17 answered all 21 items.

Internal consistency, measurement error, and test-retest reliability.

Internal consistency between MSS items was high (Table 1) for both test ($\alpha = 0.81$) and retest ($\alpha = 0.82$). Table 1 shows that Pearson's correlations coefficients were $r = 0.82$

illustrating good test-retest reliability when measures are administered 7-days apart. We found that measurement errors (limits of agreement, LOA) were acceptable (Table 1 and Figure 1) with the mean difference and SD being -1.10 ± 4.85. The Bland-Altman limits of agreement are therefore 1.10 ± 2 x 4.85 or (-8.60 to 10.80) with no participants' data falling outside two standard deviations. The t-test and sign test between test and retest show that there was not a significant mean difference between the two instances of testing (t = 0.96, p = 0.349).

<INSERT FIGURE 1 and Table 1 ABOUT HERE>

Convergent and Discriminant Validity

As the MSS and SSFS were both normally distributed and appeared linearly related, Pearson's product-moment coefficient (r) was used to examine the strength of correlation (convergence) between the MSS and the SSFS with a moderate value of r = 0.4 – 0.7 indicating that the measurement tools are related but different constructs. The correlation between the MSS initial measurement and the SSFS was r = 0.330 (p = 0.086) indicating low convergence. Pearson's product-moment coefficient (r) was also used to examine the

strength of correlation (discriminant validity) between the MSS and the RSES with a moderate value of $r < 0.2$ indicating that the measurement tools are unrelated constructs. The correlation between the MSS initial measurement and the RSES was $r = 0.154$ ($p = 0.444$). Moderate correlations were found between the Rosenberg and SSFS ($r = 0.661$ ($p < 0.001$).

Discussion

This is the first study to test the measurement properties of a scale designed to quantify the meaningfulness of a songwriting with a neurotypical university student population. Findings indicate the MSS has strong internal consistency (alpha 0.81-0.82), good test-retest reliability ($r= 0.82$), poor convergent validity ($r = 0.330$) and good discriminant validity ($r = 0.154$). We found that the LOA (measurement error) ($1.10 \pm 2 \times 4.85$) were good indicating the measures were stable. Therefore the test scores derived from the measure were reliable and valid for a neurotypical university student population.

Internal consistency and test-retest findings are consistent with Baker et al.'s (2016) study with a sample of participants in detoxification and a sample of those with acute mental illness. In this study, internal consistency was lower (here alpha 0.81-0.82 and in the

previous study 0.98, 0.96) which would reflect the more diverse spread of scores across items. It was already noted that a large majority of those with mental health challenges scored most items a 4 or 5 and this lack of spread led to a higher alpha score than in this study. It is likely that participants with mental health challenges were attempting to "please" the therapist with their positive answers given they were participants in a music therapy songwriting program. Further, all items were constructed to be positive statements potentially leading to participants answering items quickly without considerable thought. In this study, we intentionally reversed the wording of 6 of the 21 items to encourage participants to carefully read each item and consider their response. In addition, the participants in this study were students who were creating a song as part of their university training rather than being a part of a therapy process. These differences in conditions led to a greater spread in responses and a lower but acceptable level of internal consistency.

Test-retest reliability was strong in this study adding further weight to the scale's potential for use in research studies. In Baker et al.'s (2016) study, the test-retest reliability was high when the scale was administered just 6 hours after the initial testing. In this study, the measure was repeated 7 days later, and while the test-retest measures were slightly less reliable than the earlier study, they were still acceptable when testing occurred 7 days apart.

Further, in this study, the measurement error was much lower than in the two samples in Baker et al.'s (2016) study. Despite the time between testing being 7 day rather than 6 hours, this measurement stability is not surprising given these measures were administered to people with relatively stable mental health status and not so unwell that they are admitted to a hospital context.

Convergent validity in this study was substantially lower ($r = 0.330$) than in the study of participants with mental illness ($r = 0.68$; $r = 0.56$) indicating the constructs of flow and meaningfulness (as captured in the MSS) are not as related as was apparent in the Baker et al. (2016) study. The higher correlations in the earlier study are most likely a result of participants scoring relatively high for each item across both measurement scales. The lower, but still tending towards significance ($r = 0.330$, $p = 0.086$), correlation in the current study is interesting for a number of reasons. The Meaningfulness of Songwriting Scale MSS and the Short State Flow Scale (SSFS) focus on different psychological features related to the experiential aspects of the participants' immediate environment. The songwriting scale has an explicit focus upon the creative and music processes involved in constructing a new composition while the flow scale centers upon issues related to broader cognitive appraisals of the immediate situation. Thus the approaching significance correlation obtained in the measure of convergent validity suggests that these two

measurement instruments are quantifying constructs that are both different but subtlety related. The theory of flow suggests that creative activities such as songwriting may produce feelings of flow. While the results of this study are not in any way conclusive, when considered alongside the results of our previous work, there is a growing body of evidence that indicate songwriting experiences do stimulate strong flow responses.

Discriminant validity had not yet been tested in previous studies of the meaningfulness of songwriting, and in this study we chose self-esteem to be a construct we considered would not be strongly related to meaningfulness. Our hypothesis was supported here and the scales are measuring distinctly different constructs ($r = 0.154$). Interestingly, the correlations between SSFS and RSES were strong and significant ($r = 0.661$, $p < 0.001$) indicating that self-esteem and flow are more related to each other than they are to the concept of meaningfulness in a songwriting context. This might explain why flow was not as strongly correlated with MSS as previously hypothesized. It was however more strongly correlated than self-esteem with the MSS indicating that flow and meaningfulness are more related constructs than self-esteem and meaningfulness.

Despite our best efforts to recruit participants for this study (and indeed we extended the recruitment period by a semester to increase the size of the sample) and to collect data in a systematic way, there were challenges in gaining informed consent and as

detailed earlier, with missing data. Participating in this study is low on students' priorities especially when the study is being completed concurrent with the semester's assessment period. For those students presenting their song in week 12 (final class), there were no more classes after that class within which to return the 7-day retest, and as such, none of the students who performed on the last day, returned the 7-day MSS. Ethics approval was not given to "chase up" students for this missing data. A larger sample of participants and with less missing data, would have allowed for a confirmatory factor analysis to test whether the model of the meaningfulness of songwriting is captured in the items. The analysis would permit testing whether items load sufficiently on to the three different constructs of meaning – affective meaning, cognitive meaning, and relational meaning.

It would have been useful to interview the participants to explore in depth, some of the reasons why the songwriting experience and the song product was or was not meaningful to them. As all previous studies were based on working with clinical or non-clinical participants whose songwriting experiences were facilitated by a music therapist (Baker & MacDonald, 2013a/b; Baker et al., 2015; Baker et al., 2016), these songs were self-created and created as part of a university class assessment. They may not derive as

much meaning from this experience when they may be challenged to create a meaningful song that also "complies" with the songwriting criterion used to assess their songwriting skills. There is a mismatch between the songwriting context here (songwriting as a form of university assessment) and the earlier studies of Baker and colleagues (songwriting to express identity or tell one's story [Tamplin et al., 2015]), which is likely to substantially impact the degree of meaning experienced by the process and product.

When the focus is on the product either for commercial sales or in this case for university assessment, the effort the songwriters may take to create a song that ticks all of the learning objectives or includes all the components of a "hit song" are at the forefront of the songwriting experience. Therefore, a limitation of this study is that the meaning derived from the experience is likely to be qualitatively and potentially quantitatively different to a songwriting experience that resembles a therapeutic songwriting experience where the purpose is to express identity or authentically communicate one's own story. One way to tease out the impact of the "assessment" on derived meaning would be to compare the scores on MSS in a group of people whose composed songs were a component of university assessment and those where it was a non-assessable learning activity.

Conclusion

This study was designed determine the psychometric properties of The Meaningfulness of Songwriting Scale (MSS) for students (non-music majors) studying popular music songwriting. The results suggest that the measure has adequate measurement properties for use with university students and maybe utilized to further our understanding of the process and outcomes of this emerging, and potentially highly beneficial, music therapy strategy. As songwriting develops as a contemporary music therapy intervention, designed to positively influence health and wellbeing in a variety of psychological significant ways, it is crucial that we better understand the basic components of this intervention. This understanding will help in the development of models of best practice and further knowledge about tailoring the intervention to suit the needs of clients. This knowledge will also help clinicians and researchers predict the type of outcomes that can be expected. Future research should therefore seek to further investigate the psychometric properties of these instruments across different populations and contexts. In addition, to further refine theories about why songwriting experiences may be meaningful to people, qualitative interviews analysed through constructivist lenses is also needed.

References

Baker, F. A. (2015a). What about the music? Music therapists' perspectives of the role of

music in the therapeutic songwriting process. *Psychology of Music, 43*(1), 122–139. DOI: 10.1177/0305735613498919

Baker, F. A. (2015b). *Therapeutic songwriting: Developments in theory, methods, and practice*. London: Palgrave Macmillan.

Baker, F. A. (2016). Therapeutic Songwriting: Future Directions. In C. Dileo (Ed.), *Envisioning the future of music therapy* (pp.123–130)*,* Philadelphia, PA: Quality of Life Research Centre.

Baker, F. A., & MacDonald, R. A. R. (2013a). Flow, identity, achievement, satisfaction and ownership during therapeutic songwriting experiences with university students and retirees. *Musicae Scientiae, 17*(2), 129–144. doi: 10.1177/1029864913476287.

Baker, F. A., & MacDonald, R. A. R. (2013b). Experiences of creating personally meaningful songs within a therapeutic context. *Arts & Health*, 1–19. doi: 10.1080/17533015.2013.808254.

Baker, F. A., Rickard, N., Tamplin, J., & Roddy, C. (2015). Flow and meaningfulness as mechanisms of change in self-concept and wellbeing following a songwriting intervention for people in the early phase of neurorehabiltation. *Frontiers in Human Neuroscience, 9,* 299. DOI: 10.3389/fnhum.2015.00299

Baker, F. A., Silverman, M. J., & MacDonald, R. A. R. (2016). Reliability and validity of

the Meaningfulness of Songwriting Scale (MSS) with adults on acute psychiatric and detoxification units. *Journal of Music Therapy*, *53*(1), 55–74. doi:10.1093/jmt/thv020

Baker, F., Wigram, T., Stott, D., & McFerran, K. (2008). Therapeutic songwriting in music therapy: Part 1. Who are the therapists, who are the clients, and why is songwriting used? *Nordic Journal of Music Therapy, 17*, 2, 105–123. DOI:10.1080/08098130809478203

Baker, F., Wigram, T., Stott, D., & McFerran, K. (2009). Therapeutic songwriting in music therapy: Comparing the literature with practice across diverse populations. *Nordic Journal of Music Therapy, 18,* 1, 32–56. DOI: 10.1080/08098130802496373

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, i,* 307–310.

Blascovich J., & Tomaka J. (1991). Measure of self-esteem. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds), *Measures of personality and social psychological attitudes* (pp. 115–160). San Diego, CA: Academic Press.

Choi, A., Lee, M. S., & Lee, J. (2010). Group music intervention reduces aggression and improves self-esteem in children with highly aggressive behavior: A pilot controlled trial. *Evidence-Based Complementary & Alternative Medicine (Ecam)*,

7(2), 213. doi:10.1093/ecam/nem182

Clark, I. N., Baker, F. A., Peiris, C. L., Shoebridge, G., & Taylor, N. F. (2015). The Brunel Music Rating Inventory-2 is a reliable and valid instrument for older adults in cardiac rehabilitation selecting music for exercise. *Psychology of Music, 44*(2), 249–262. DOI: 10.1177/0305735614565830.

Grocke, D., Bloch, S., & Castle, D. (2009). The effect of group music therapy on quality of life for participants living with a severe and enduring mental illness. *Journal of Music Therapy, 46*(2), 90–104. doi:10.1093/jmt/46.2.90

Hald, S. V., Baker, F. A., Ridder, H. M. O. (2015). A preliminary psychometric evaluation of the Interpersonal Communication Competence Scale for acquired brain injury. *Brain Injury, 29*(9), 1105–1112. DOI:10.3109/02699052.2015.1024740

Imms, W., Jeanneret, N., & Stevens-Ballenger, J. (2011). *Partnerships between schools and the professional arts sector: Evaluation of impact on student outcomes*. Melbourne: Arts Victoria.

Jackson, S., & Csikszentmihalyi, M. (1999). *Flow in sports: The keys to optimal experiences and performances.* Champaign, IL: Human Kinetics Book.

Magee, W. L., Siegert, R. J., Taylor, S. M., Daveson, B. A., & Lenton-Smith, G. (2016). Music Therapy Assessment Tool for Awareness in Disorders of Consciousness

(MATADOC): Reliability and validity of a measure to assess awareness in patients with disorders of consciousness. *Journal of Music Therapy*, *53*(1), 1–26. DOI:10.1093/jmt/thv017

Martin, A. J., & Jackson, S. A. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: Examining 'short' and 'core' flow in diverse performance domains. *Motivation and Emotion, 32*, 141–157. DOI: 10.1007/s11031-008-9094-0

O'Brien, E. (2014). *The effect and experience of therapeutic songwriting on adult cancer patients' quality of life, mood, distress levels and satisfaction with hospital stay*. Doctoral Dissertation. Melbourne: The University of Melbourne

Porter, S., Holmes, V., McLaughlin, K., Lynn, F., Cardwell, C., Braiden, H., & ... Rogan, S. (2012). Music in mind, a randomized controlled trial of music therapy for young people with behavioural and emotional problems: study protocol. *Journal of Advanced Nursing*, *68*(10), 2349–2358. doi:10.1111/j.1365-2648.2011.05936.x

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice*. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Robb, S. L., Burns, D. S., Stegenga, K. A., Haut, P. R., Monahan, P. O., Meza, J., & ... Haase, J. E. (2014). Randomized clinical trial of therapeutic music video intervention

for resilience outcomes in adolescents/young adults undergoing hematopoietic stem cell transplant: A report from the Children's Oncology Group. *Cancer*, (6), 909–917. DOI: 10.1002/cncr.28355

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Silverman, M. J. (2011). Effects of music therapy on change and depression on clients in detoxification. *Journal of Addictions Nursing, 22,* 185–192. doi:10.3109/10884602.2011.616606

Silverman, M. J. (2012). Effects of group songwriting on motivation and readiness for treatment on patients in detoxification: A randomized wait-list effectiveness study. *Journal of Music Therapy, 49,* 414–429.

Silverman, M. J., Baker, F. A., & MacDonald, R. A. R. (2016). Flow and meaningfulness as predictors of therapeutic outcome within songwriting interventions: Two post hoc analyses. *Psychology of Music*. First published February 5th 2016. DOI: 10.1177/0305735615627505

Tamplin, J., Baker, F.A., Rickard, N., Roddy, C, & MacDonald, R. (2015). A theoretical framework and therapeutic songwriting protocol to promote integration of self-concept in people with acquired neurological injuries. *Nordic Journal of Music Therapy, 25*(2), 111–133. Doi.10.1080/08098131.2015.1011208.

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test–retest reliability. *PLoS ONE*, *8*(9), e73990. http://doi.org/10.1371/journal.pone.0073990

Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. [Research Support, Non-U.S. Gov't]. *Statistics in medicine, 17*(1), 101–110. doi: 10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E

Wood, L., Ivery, P., Donovan, R., & Lambin, E. (2013). 'To the beat of a different drum': improving the social and mental wellbeing of at-risk young people through drumming. Discovering relationships using music - beliefs, emotions, attitudes, and thoughts. *Journal of Public Mental Health*, *12*(2), 70–79. doi:10.1108/JPMH-09-2012-0002.

Table 1 *Reliability of the MSS*

Figure 1. Bland-Altman plot: Difference against participants' test and retest mean for MSS data